

Running Head: Cross-age Peer Tutoring and Fluency-based Instruction

Cross-age Peer Tutoring and Fluency-based Instruction to Achieve Fluency with
Mathematics Computation Skills: A Randomized-Controlled Trial

Irene Greene, MSc¹

Aoife Mc Tiernan, PhD²

Jennifer Holloway, PhD¹

Conflict of Interest: The authors declare that they have no conflict of interest.

¹School of Psychology

National University of Ireland, Galway, Ireland

²School of Psychology

University of South Wales

Corresponding author: Aoife Mc Tiernan, PhD, School of Psychology, University of South Wales. Email: aoife.mctiernan@southwales.ac.uk

Abstract

The current study employed a randomized controlled trial (RCT) to evaluate the use of peer tutoring and fluency-based instruction to increase mathematics fluency with addition and subtraction computation skills. Forty-one elementary school students between the ages of eight and 12 years participated in the eight-week study using cross-age peer tutoring, Say All Fast Minute Every Day Shuffled (SAFMEDS), frequency-building, and the Morningside Math Facts curriculum (Johnson, 2008). Pre- and post-test measures of mathematics fluency and calculation were conducted with all participants. A measure of social skills and competing problem behaviors was also conducted at pre- and post-testing to evaluate any additional effects of the peer-tutoring model. The results demonstrated a statistically significant difference between groups on measures of mathematics fluency, with the Experimental Group demonstrating significantly higher scores than the Control Group at post-testing. There were no significant differences between groups on measures of social skills and competing problem behaviors or calculation. The findings indicate that cross-age peer-tutoring and fluency-based instruction resulted in positive outcomes for tutees in the mathematics domain, specifically mathematics fluency.

Keywords: peer tutoring; behavioral fluency; SAFMEDS; frequency-building; mathematics

Cross-age Peer Tutoring and Fluency-based Instruction to Achieve Fluency with Mathematics Computation Skills: A Randomized-Controlled Trial

Proficiency in mathematics is not only necessary for academic success, but is essential for daily living (Geary, Hoard, Nugent, & Bailey, 2012). The application of basic computational skills is required for independent adaptive functioning and creates a foundation for time, money management, and problem solving (Patton, Cronin, Bassett, & Koppel, 1997). Rivera-Batiz (1992) found that mathematic competence was a major factor in accounting for higher levels of employment, work productivity, and income even after the impact of IQ and reading achievement as explanatory variables had been considered. Mathematical ability can be likened to basic literacy in that it is a vital component for independent functioning within society (Geary, 2013).

Despite the acknowledged importance of mathematic skills, deficits in this area continue to be observed within school systems, with 18% of fourth grade students in the U.S. performing below the basic achievement level (Kena et al., 2016). In Ireland, similar difficulties with mathematics have been reported. The Counting on Success Report, prepared by the Irish Department of Education, found that 15% of Irish fourth grade students were found to be at or below the minimum proficiency level for mathematics (Surgenor, Shiel, Close, & Millar, 2006). The data compiled in this report were based upon tests developed to be compatible with the framework of the national school curriculum for mathematics. More recently, the National Assessment of Reading and Mathematics (Shiel, Kavanagh, & Millar, 2014) found that 6.2% of students in second grade and 5% in sixth grade were performing below the lowest proficiency level. These deficits increase when looking at data from urban disadvantaged schools. Within urban schools classified as having the highest disadvantage, 14.3% of second grade students and 11.9% of sixth grade students performed below then lowest proficiency level measured (Shiel et al., 2014).

Socio-economic status (SES) has been identified as an influential variable impacting mathematics proficiency. The rate of students performing below proficiency levels almost doubles in populations living in urban poverty (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003) and students from high-poverty backgrounds in particular demonstrate large deficits in mathematics thinking and skills (Griffin, Case, & Siegler, 1994). Findings from the Program for International Assessment (PISA) demonstrated a positive correlation between SES and mathematics performance (OECD, 2016). For students from the U.S., 13.1% of the variance in mathematics scores was found to be attributable to SES. Since family poverty, parental education, low grades, and low test scores are all major predictors of school failure and drop-out (Thompson & Vance, 2001), research focusing on mathematic instruction amongst disadvantaged populations is critical.

A fundamental skill in the progression of mathematic abilities is the computation of math facts. In order to demonstrate proficiency in mathematics and to progress to more complex mathematics skills, students should achieve mastery of key mathematic concepts (Nelson, Burns, Kanive, & Ysseldyke, 2013). Further, those who are both accurate *and* fluent when performing basic arithmetic have been found to perform better on assessments of overall mathematics ability (Carr, Steiner, Kyser, & Biddlecomb, 2008). Research by Jordan, Hanich, and Kaplan (2003b) found that children with poor arithmetic fact mastery (defined as the ability to compute an addition or subtraction number fact in 3 seconds or less) in second and third grade performed significantly lower on tests of broad mathematics achievement and showed very little growth in timed number facts when compared to those with good fact mastery.

Fluent responding can be defined as accurate and quick responding to a selected stimulus, with minimal effort, that enables an individual to function efficiently and effectively in their natural environment (Axtell, McCallum, Bell, & Poncy, 2009; Binder,

1996; Cates & Rhymer, 2003). When students achieve a fluent level of performance (i.e., a higher rate of correct responding) with such skills, they typically retain and maintain what they have learned, remain on task in the face of distraction and for longer periods, and more readily apply what they have learned to new situations (Binder, 1996; Brady & Kubina, 2010; Johnson & Street, 2013; Kubina & Wolfe, 2005). It has been suggested that students who are slow but accurate in responding to math facts may not be able to complete assigned tasks within set time limits and consequently may receive less or no reinforcement (Bliss et al. 2010; Skinner, 2002). A lack of fluency in basic computations can lead to mathematics difficulties that can persist across the lifespan (Nelson et al., 2013). Addressing these deficits early provides a greater chance for success, before the school mathematics curriculum increases in scope and difficulty (Fuchs, Fuchs, Yazdian, & Powell, 2002).

Fluency-based instruction (i.e., instruction designed to increase rate of correct responding with targeted skills) leads to responses with high accuracy and quick response times that require seemingly low effort for the responder (Weiss, Pearson, Foley, & Pahl, 2010). Research into interventions that promote fluency has found practice to be a key active ingredient (Burns, 2005; Coddling, Burns, & Lukito, 2011; Daly, Martens, Barnett, Witt, & Olson, 2007). In addition to practice, successful interventions utilize explicit instruction, drill, feedback, and reinforcement (Burns, 2005; Coddling et al., 2011; Fuchs, Fuchs, Powell, Seethaler, Cirino, & Fletcher, 2008). A number of fluency-based instructional approaches have been developed, and demonstrated as effective, to increase fluency with mathematics skills. Three of these which incorporate the components mentioned above include explicit timings (ET; Van Houten & Thompson, 1976), SAFMEDS (Casey, McLaughlin, Weber, & Everson, 2002), and peer tutoring (Hawkins, Musti-Rao, Hughes, Berry, & McGuire, 2009).

Explicit timing (ET) is a method of fluency based instruction that uses a timing procedure to increase the speed with which a person responds to chosen stimuli. It involves

the presentation of a task that can be accurately completed and a finite amount of time in which it is to be completed (Schutte et al., 2015). It provides repeated practice for mathematics fluency, with students completing as many repetitions of the target behavior within the timing, usually one minute. Research to date has shown ET to be an effective procedure for increasing rate of academic responding, and is particularly effective when consequence procedures such as feedback, self-correction, and goal setting are incorporated (Gross et al., 2014; Miller, Hall, & Heward, 1995). Frequency-building procedures combine timed repetition of performance with performance feedback to increase fluency with component skills (Kubina, Yurich, Durica, & Healy, 2015). A number of studies to date have demonstrated the ability of frequency-building to increase rate of accurate responding within the mathematics domain (Gallagher, 2006; Hartnedy, Mozzoni, & Fahoum, 2005; McTiernan, Holloway, Healy, & Hogan, 2016; Poncy, Duhon, Lee, & Key, 2010).

Say All Fast Minute Each Day Shuffled (SAFMEDS) is a frequency-building intervention (Kubina, Yurich, Durica, & Healy, 2016) developed to enable students to learn and increase rate of correct responding with key facts (Graf & Lindsley, 2002). SAFMEDS flashcards are used within timed practice sessions to engage in free operant responding and repetitions of the target behavior. Flashcards are developed dependent on the behavior targeted for increase and on the skill level of the student. Performance feedback can also be provided using the flashcards subsequent to timed practice. To date, positive outcomes have been demonstrated in the literature to support its use to increase fluency in the mathematics domain (Beverly, Hughes, & Hastings, 2009; Casey et al., 2003; Chapman, Ewing, & Mozzini, 2005; Hartnedy et al., 2005; Hunter, Beverley, Parkinson, & Hughes, 2016). Although a number of studies have demonstrated positive outcomes as a result of fluency-based instruction, a meta-analysis of mathematics interventions conducted by Coddington, Hilt-Panahon, Panahon, and Benson (2009) found that, while several fluency-based instructional

approaches produced positive results, many of those included failed to meet criteria to be considered evidence-based.

Peer tutoring is an instructional method that has the benefits of individualising content based on students' needs, while allowing for extensive feedback and maintaining high levels of task engagement (Dufrene, Noell, Gilbertson, & Duhon, 2005). It is a student mediated instructional procedure, where one student is responsible for providing instruction intended to teach another student in a specific skill area (Dufrene et al., 2005; Robinson, Schofield, & Steers-Wentzell, 2005). Research indicates that students benefit from engaging in the role of both tutor and tutee (Robinson et al., 2005). The process of explaining concepts and facts is thought to support tutors to develop and reinforce their own skills (Mitchell, Morrison, Feinauer, Wilcox, & Black, 2015).

A diverse literature base exists for peer tutoring across a variety of content areas, intervention intensities, participant ages and ability levels (Robinson et al., 2005). A number of reviews examining this instructional approach have found that it has beneficial outcomes; namely, improved academic outcomes across a variety of content areas, as well as having collateral effects on a variety of behavioral and social outcomes (Bowman-Perrott et al., 2013; Bowman-Perrott, Burke, Zhang, & Zaini, 2014; Robinson et al., 2005; Stenhoff & Lignusaris-Kraft, 2007). Bowman-Perrott et al. (2014) conducted a meta-analysis of single-case research, examining both the academic benefits and the social and behavioral outcomes of peer tutoring. They found that in addition to the primary academic benefits, peer tutoring led to increases in positive social interactions, academic engagement, and decreased disruptive and off-task behaviors. A meta-analysis by Robinson et al. (2005), also found increases in on-task behaviour, attendance, positive attitudes about school, self-rated scholastic competence and behavioural conduct, self-concept, and sense of belonging. These observed benefits are not limited to the tutees, but have also been found in tutors across

studies. Specific to mathematics instruction, peer tutoring has been demonstrated as an effective instructional approach to teach mathematics skills (Fueyo & Bushell, 1998; Sprinthall & Scott, 1989).

A number of variations of peer tutoring exist, for example, heterogeneous grouping (in which tutees are taught by peers from the same grade level with a higher level of knowledge or skill), homogeneous grouping (in which tutees are taught by peers with similar levels of knowledge or skill), and cross-age peer tutoring. Cross-age peer tutoring involves an older tutor who teaches a younger tutee (Stenhoff, & Lignusaris-Kraft, 2007). While cross-age peer tutoring has been found to be an effective intervention for both academic and collateral gains, there are far fewer studies of cross-age tutoring than of same-age peer tutoring (Robinson et al., 2005). Previous research on peer tutoring at an elementary school level has primarily focused on using tutors from the same grade level (Ginsburg-Block, Rohrbeck, & Fantuzzo, 2006; Rohrbeck et al., 2003) and much of the research on cross-age peer tutoring has concentrated on students with learning disabilities and emotional and behavioral disorders (Heron, Welsch, & Goddard, 2003). Additional research is necessary to evaluate cross-age peer tutoring and to investigate the effects of this instructional approach in general education classes (Stenhoff & Lignusaris-Kraft, 2007).

Purpose of the Present Study

The purpose of the current study was to extend the literature on both cross-age peer tutoring and fluency-based instructional approaches to increase fluent performances with mathematics skills, specifically with students attending classrooms situated in disadvantaged areas. Findings that SES influences mathematics proficiency (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003) indicate that students attending such classrooms may benefit from additional instruction in this domain. Peer-tutoring was chosen as an instructional approach

since it can be used to increase students engaged time, in particular to build fluency with academic skills (Stenhoff & Lignugaris-Kraft, 2007), and can be implemented with relatively low demands on resources (Dufrene et al., 2010). Cross-age peer tutoring was evaluated in combination with explicit timings using SAFMEDS. Kubina, Yurich, Durica, and Healy (2016) posit that building frequency to a performance criterion results in systematically increasing the frequency of the target skill, and that cumulative implementations of such frequency-building sessions foster growth across time. Therefore, fact families from the Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008) and associated performance criterion were targeted for intervention within the current study.

A stratified randomized controlled trial (RCT) was employed to evaluate the effectiveness of the intervention in comparison to a treatment as usual (TAU) control condition. The effects of the intervention were evaluated using standardised norm-referenced assessments of mathematics ability, direct measures of fluency with targeted math facts, and a measure of social skills and competing problem behaviors. Each measure was recorded pre- and post-intervention to evaluate if the intervention group would demonstrate significantly greater performances on all outcome measures at post-testing. Since previous research investigating peer-tutoring has indicated benefits from engaging in both the role of the tutor and tutee (Robinson et al., 2005), each measure was also conducted with the tutoring group to investigate possible increases in outcome measures associated with the tutor role alone in non-reciprocal peer-tutoring.

Method

Participants and Setting

Forty-four participants were recruited from third, fourth, and fifth grade in an all-female urban elementary school in Galway, Ireland. An a priori power analysis was carried

out to determine a sufficient sample size. Using an alpha of .05 and a power of .95, and a medium effect size ($f^2 = .25$; Faul, Erdfelder, Lang, & Buchner, 2007) the desired sample size was determined to be a total of 35 participants. Due to attrition, three participants did not complete the intervention resulting in a final sample of 41 participants. Causes of attrition included participants relocating to another school or extended absences from school while the study was taking place. Participants were between the ages of eight and 12 years ($M = 9.94$, $SD = 0.96$). The participating school was recruited through convenience sampling. Once teachers from each grade agreed for their students to participate, appropriate consent and assent forms for the parents and students were distributed by the experimenter and were returned by the participants and their parents prior to pre-testing.

The study was conducted in the participants' school, a Band One school within the School Support Programme (Delivering Equality of Opportunity in Schools; DEIS) by the Irish Department of Education (DES, 2005). DEIS schools are identified as those with the highest level of disadvantage, who receive additional supports and resources. Band One schools are those where the level of disadvantage is greatest. DEIS urban primary schools report higher levels of pupils with literacy and numeracy difficulties as well as emotional and behavioral difficulties. Absenteeism is also reported as a more significant problem in DEIS urban band 1 schools (Smyth & McCoy, 2009). The school had mixed-sex classes up to second grade and all-female classes from there upwards. The study took place in a room separate to the participants' regular classroom, either prior to break time in the morning or after lunch in the afternoon. Participants came to the room in groups of 5 dyads, and each dyad sat across the table from each other. The experimenter was in the room throughout each session, and was aided by a second instructor during several sessions.

Intervention Agents and Training

Both the experimenter and an additional instructor were Masters level students completing university postgraduate training in Applied Behavior Analysis. The experimenter provided training sessions in frequency-building and peer-tutoring intervention to the second instructor prior to and during the course of the intervention.

Materials

Flashcards. Sets of flashcards consisting of targeted math facts were developed for frequency-building using SAFMEDS. Facts were printed on the front of a flashcard and the answer was printed on the reverse. The tutee was required to *see* the math fact on the front of the flashcard, and then to *say aloud* the answer that was printed on the reverse of the card. Math facts were printed on laminated rectangular cards, measuring 10 x 5cm. Facts were printed in black font on a white background. The answers on the back of the cards were printed in grey to prevent the answer being visible to the tutee. Each set of tutoring cards was contained in an envelope and labelled to indicate the math facts set inside. Each set contained 70 cards, as this was 10 more than the fluency aim. Further, based upon pre-test rates of responding, the experimenters anticipated that participants would not score higher during timings. Cumulative sets of previously mastered targets contained 80 cards, as the fluency aim for these was 60-70. A green 'Go' was placed on the front of each envelope to indicate the current set of problems that the tutee was working on.

Curriculum. The Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008) was used to identify component skills in mathematics which were targeted during intervention. Within this curriculum, math facts are presented in fact families which allows the learner to create four facts from a three-number family (for example, by memorising the 2,3,5 family, students combine these numbers to get $2+3=5$, $3+2=5$, $5-3=2$, $5-$

2=3). The addition and subtraction curriculum is broken down into 16 levels or slices, each containing between three and four fact families. For the current intervention, each set of flashcards was based on a slice of the curriculum containing the fact families within that slice. There were four cumulative sets interspersed throughout the intervention consisting of a mixture of math facts from previously mastered sets.

Fluency aims used in the current study were based on those within the Morningside curriculum. Participants were required to reach a fluency aim of between 50-60 correct responses per minute for each set of flashcards containing newly introduced math facts, and a fluency aim of 60-70 correct responses per minute for cumulative sets. At the beginning of the intervention, all participants were assessed on the first set of flashcards, which was based on the first slice of the curriculum. No participant was found to perform fluently; therefore, all participants began with this introductory set. They then progressed, in sequence, through flashcard sets based on each subsequent slice of the curriculum. Progression to the next set was contingent on achieving the fluency aim on the previous set.

Tutoring folders. Tutoring folders were made from manila file folders. On the inside left, there was a progress chart consisting of 10 steps depicted by either ten flowers or ten stars for the tutees to colour in. Tutees coloured in a step each time they beat their previous timing by five or more correct responses. Participants had a chart in their folders on which they received checks as positive reinforcement during the tutoring sessions. A large emoticon and a large 'X' were placed on the back of the folder to record correct and incorrect responses to math facts during timings (i.e., correct responses to the presentation a flashcard were placed on the emoticon and incorrect responses placed on the 'X'). Each dyad was also provided with a data collection sheet where the tutees progress was recorded. Data sheets were divided into sections pertaining to each slice of the curriculum with rows representing an intervention session and columns used to record correct and incorrect responses after each

timing. Either the tutor or tutee wrote in the number of correct and incorrect math facts after each timing.

Timer. A small digital timer with a countdown setting was used to keep time during the testing sessions. Each dyad had their own individual timer.

Dependent Measures

To evaluate outcomes of the intervention on mathematics skills, two subtests of the Woodcock Johnson Tests of Achievement, Third Edition (WJIII; Woodcock, McGrew, & Mather, 2001b); namely, Mathematics Fluency and Calculation, were conducted with all participants at pre- and post-testing. A direct measure of fluency with math facts targeted during intervention was also tested pre- and post-intervention. To investigate any impact on social skills, the Social Skills Improvement System (Gresham & Elliott, 2008) was also completed at pre- and post-testing with all participants.

Woodcock Johnson Test of Achievement, Third Edition. The WJIII (Woodcock, McGrew, & Mather, 2001b) is an individually administered assessment for measuring skills in reading, mathematics, writing, oral language abilities, and academic knowledge. The internal consistency of the WJIII (Woodcock, McGrew, & Mather, 2001b) has been reported at .98 (McGrew & Woodcock, 2001) and correlation coefficients of between .65 and .79 have been reported between the WJIII (Woodcock, McGrew, & Mather, 2001b) and other leading measures of achievement (McGrew & Woodcock, 2001). Two mathematics subtests of the WJIII (Woodcock, McGrew, & Mather, 2001b) were used in the study. These subtests measure computational skills and automaticity with basic math facts, providing a measure of basic math skills. Test-retest reliabilities for the calculation and mathematics fluency subtests of the WJIII have been reported at .86 and .90, respectively (Schrack, F. A., McGrew, K. S., & Woodcock, R. W., 2001).

Calculation. This subtest measures mathematics achievement in relation to ability to access and apply knowledge of numbers and calculation procedures. It is a paper and pencil test involving arithmetic and computation. Participants are presented with the problems on a worksheet and are asked to provide written answers on the sheet next to each problem.

Mathematics Fluency. This subtest measures mathematics achievement; specifically, the ability to access and apply automatically, and with speed, digit-symbol arithmetic procedures. Participants are presented with a sheet of simple calculations and asked to answer as many as they can correctly in three minutes. The subtest includes two sheets in total with 80 calculations on each.

Fluency with target math facts. Flashcards consisting of math facts from the Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008) were used to directly assess fluency with math facts targeted across intervention sets. Each participant was assessed using multiple exemplars of addition and subtraction problems targeted during intervention (Mc Tiernan et al., 2016). Seventy problems relevant to the first 7 fact families in the Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008) were presented in random order. Participants were assessed on the same problem set at pre- and post-testing. Facts were printed on the front and answers printed on the reverse of the flashcards. As many cards as possible are presented in a one minute timing, after which the number of correct and incorrect responses are counted and recorded. The assessment procedure required the experimenter to show the participant the front of the tutoring card and the participant to answer the math fact. The experimenter waited 3 seconds for a response. If an incorrect response was given or there was no response, the experimenter moved to the next card and an incorrect response was recorded. If the participant gave a correct response within 3 seconds, the experimenter moved to the next card and a correct response was recorded.

Social Skills Improvement System-Rating Scales. The Social Skills Improvement System Rating Scales (SSIS-RS; Gresham & Elliott, 2008) is a multi-rater series of rating scales that includes rating scales for teachers, parents, and participants. It is designed for children aged 8-18 years and is a revised form of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990). It documents the frequency of social skills and competing problem behaviors. Forms include common social skills in the following subdomains: communication, assertion, responsibility, empathy, engagement, and self-control. The Problem Behavior subscales include the subdomains of: externalizing, bullying, hyperactivity/inattention, internalizing, and autism spectrum. Participants indicate how true a statement is about each social skill and problem behavior for them using a 4-point scale of *not true, a little true, a lot true* and *very true*. The scales are reported to have a high reliability from a national standardization sample (N = 4700) reported in the SSIS-RS manual (Gresham & Elliott, 2008). The reliability of subscales is between .70-.80 for the student forms and test-retest reliability is between .50-.80. In the current study, the Student Response form for ages 8-12 years was completed by each participant.

Independent Variables

Control condition. At the beginning of the intervention there were 15 participants in the control condition; however, due to attrition, only data from 14 participants were included in the analysis. Participants in the Control Group were in third or fourth grade and participated in treatment as usual (TAU) in the form of typical classroom mathematics instruction. They did not receive any component of the intervention (i.e., peer-tutoring or frequency-building using SAFMEDS and targets from the Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008). The intervention sessions took place outside of scheduled classroom mathematics instruction, so during intervention times the control group remained in their regular classroom and participated in other subject instruction.

Experimental condition. There were 15 participants from third and fourth grade in the experimental condition. The Experimental Group also received TAU in the form of typical classroom mathematics instruction, the same as the Control Group. In addition to this, they participated as tutees in the peer-tutoring intervention for 30 minutes, approximately 3 days per week, for 8 weeks. Participants in this group were tutored by peers in the fifth grade who were trained to implement frequency-building using SAFMEDS and targets from the Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008) with their peers. These intervention sessions took place outside of scheduled classroom mathematics instruction and so were in addition to TAU.

Tutor condition. Fourteen participants began in the tutor condition but due to attrition this number decreased to 12 during the intervention. Participants in the Tutor Group were fifth grade students and received TAU in the form of typical classroom mathematics instruction. In addition to this, they participated in tutor training, followed by tutoring their third and fourth grade peers in the Experimental Group for 30 minutes, approximately 3 days per week, for 8 weeks. These intervention sessions took place outside of scheduled classroom mathematics instruction and so were in addition to TAU.

Procedure

A stratified Randomized Control Trial (RCT) was used to evaluate the efficacy of the intervention. Tutees (those in third and fourth grade) were stratified into pairs according to their pre-test scores on the WJIII (Woodcock, McGrew, & Mather, 2001b) Mathematics Fluency subtest. Participants were matched into pairs according to the participant who had the next closest standard score to theirs. Participants in and each pair were then randomly assigned to either the Control ($n = 14$) or Experimental Group ($n = 15$) using stratified randomization design. Pre-test dependent measures were recorded for all participants in the

Experimental, Control, and Tutor Groups. Subsequent to this, the Experimental and Tutor Groups participated in the peer-tutoring intervention while all groups continued to receive TAU. The intervention was carried out over 8 weeks, with an additional 4 weeks for pre- and post-testing. Once the intervention was complete, dependent measures were again conducted with each participant.

Intervention

Tutor training. Tutors participating in the intervention were trained by the experimenter prior to the beginning of the intervention. Training was conducted in a separate session with only the participants who were acting as tutors. Tutors were trained in two groups, with eight tutors in the first group and six in the second group. Training sessions lasted 30 minutes each, during which tutors were instructed in the steps to be carried out during the tutoring procedure. Tutors were given a set of materials (folders, flashcards, and recording sheets) the same as those used during the intervention. The experimenter followed a training script (Appendix 1) for these sessions, adapted from Nobel (2005), which categorised the training procedure into segments: i) introduction to tutoring, ii) tutoring procedure, iii) recording.

The training followed a model-lead-test instructional format that provided opportunities for active student responding during training. This instructional strategy involved the experimenter modelling each part of the tutoring procedure for the tutors, leading them through the procedure as a group, and finally assessing the tutors' independent use of the procedure in dyads. The experimenter demonstrated the tutoring activities to the tutors and then the tutors engaged in role-playing of the activities in pairs, with one acting as tutor and the other as the tutee. During the training session, emphasis was placed on the presentation of the math facts, the correction procedure, and recording of responses. At the

end of the training session, there was a 'test' portion to determine if tutors could accurately carry out the intervention procedure. During the 'test' portion of the training, procedural integrity was recorded for each of the tutors. Procedural integrity checklists were used allowing the experimenter to calculate the number of steps implemented correctly by each tutor. Training was only considered complete when the students demonstrated the tutoring procedures with 90% accuracy. All the tutors demonstrated the procedures with at least 90% accuracy during the training session, with an average of 95.48% accuracy.

Tutoring sessions. Tutoring sessions were conducted on average three days per week for 30 minutes. Approximately 15 minutes were spent on frequency-building with SAFMEDS, with a mean of 5.21 timings per session for each tutee. The intervention lasted a total of 8 weeks. Participants completed an average of 15.3 intervention sessions, equating to 229.5 minutes of total intervention time. This included time for participants to come to the tutoring room and collect their materials and time at the end to tidy up their materials and return to their classrooms. The participants were divided into peer-tutoring dyads, with a fifth grade participant acting as the tutor and a third or fourth grade participant acting as the tutee. Outside of this requirement, the dyad assignments were random. Tutoring dyads rotated each week. Due to the uneven number of participants in the tutor and tutee groups, some tutors were required to participate in two sessions each day. The selection of these tutors was random and changed each day.

In the first session, the experimenter explained the tutoring procedure to both the tutors and tutees and described how each session thereafter would be carried out. At the beginning of the session, tutees collected their tutoring folder from the front of the classroom. Each assigned tutor and tutee pair then sat at their table facing their partner. Once all participants were seated in their pairs, the experimenter directed the participants to take their

recording sheets and reinforcement sheets from their folders, and to get their targeted flashcards out of their “Go” envelope.

Tutees were told that they needed to achieve a pre-determined fluency aim (e.g., between 50-60 correct responses per minute) to move onto the subsequent set of flashcards and that there were a total of 16 fluency aims corresponding to each set. Tutees could perform anywhere within the provided fluency range. At four intervals in the curriculum (after slice 7, 10, 13 and 16) tutees worked on a cumulative set of flashcards from the previously mastered sets. The fluency aim set for this was 60-70 correct responses per minute. Tutoring dyads conducted frequency-building using SAFMEDS throughout each session.

Frequency-building using SAFMEDS. Tutors set their timers to one minute before starting each timing. Tutors first shuffled the set of cards, began the timer, and then presented one card at a time to the tutee. The tutee answered the math fact presented on the tutoring card. If the tutee gave the correct response for a card, that card was placed on the emoticon printed on the back of the tutoring folder. If the tutee responded incorrectly, or did not give a response within 3 seconds, the card was placed on the X on the back of the folder. The tutor presented as many cards as possible in the one-minute timing. No immediate feedback was provided for responses during the timings. Feedback on responses was provided at the end of each timing.

Corrective feedback. When the one minute timing elapsed, the tutor represented the cards that were placed on the incorrect pile by holding up the cards for the tutee in the same way they were presented during the timing. If the tutee answered correctly the tutor said “Well done!” and moved to the next card in the pile. If the tutee gave an incorrect answer, or did not respond within 3 seconds, the tutor provided an error correction procedure using

least-to most prompts. This consisted of first telling the tutee to “Try again”. If a correct response was then given, the tutor would say “Well done!” and move on to the next card. If the tutee again provided an incorrect response or no response, the tutor provided the answer and asked the tutee to repeat the answer. This was completed for each card answered incorrectly during the timing. Once tutors had carried out the error correction procedure, they reset the timer and repeated the procedure.

Reward contingencies. It was explained to participants in the first session that they would receive “checks” based on four target behaviors. The first target behavior was “working well” and checks were delivered contingent on correct responding and working independently. The second target behavior was “working fast” and checks were delivered if participants beat their score on their previous timing. The third target behavior was “getting your goal” and participants received checks in this category for achieving a fluency aim range (e.g., between 50 and 60 correct responses) or performing above this criterion. The final target behavior was “listening” and participants received checks when they followed instructions (e.g., “take out your recording sheets”) throughout the session. Each participant had a chart on the desk beside their folder with pictures and words representing each target behavior.

The experimenter provided rewards for target behaviors throughout the sessions by placing a check on the relevant category and specifying why they had received the check (e.g., “you got your goal!”). Checks were delivered on a fixed ratio schedule for “working fast” and “getting your goal”. Each time one of these behaviors occurred a check was delivered. Checks for “working well” and “listening” were delivered on a variable ratio schedule. Once four checks were received, the participant was awarded a raffle ticket. Alternatively, one check on the “getting your goal” category was immediately rewarded with a raffle ticket, ensuring a higher magnitude of reward for achieving a fluency aim. At the end

of each week, a raffle was held and one participant won a prize (e.g., colouring pencils, pens, hair clips, and small toys). Tutees also had progress charts, which consisted of 10 steps, depicted by flowers or stars. Each time a tutee beat their previous timing by five or more correct answers they were instructed to colour in a step. This provided additional opportunities for reinforcement for tutees for increases in rate of correct responding.

Data collection. After each timing was carried out, either the tutors or tutees recorded the number of correct and incorrect responses on their recording chart. Progression to the next set of flashcards was contingent on the attainment of each fluency aim. When a fluency aim was not attained, additional timings were subsequently completed by the tutor and tutee on the same set. The data from each session was transferred to a graphic presentation of correct and incorrect responses for each timing by the experimenter, using Standard Celeration Charts. This provided a standard display of frequency and allowed the experimenter to assess each tutee's growth of learning across time.

Interobserver Agreement

Interobserver agreement (IOA) was conducted for 14.63% of the pre-test fluency timed probes and 17.07% of the post-test fluency timed probes. Agreement was defined as scoring the same response for each observation. Percentage agreement was calculated by dividing the number of agreements on responses correct plus agreements on responses incorrect by the number of agreements on responses correct and agreements on responses incorrect plus disagreements on responses correct and disagreements on responses incorrect and multiplying by 100. Agreement between the experimenter and second instructor was calculated resulting in 100% agreement for pre-test probes and 98.65% for post-test probes.

Procedural Integrity

The experimenter was present for every session and monitored implementation using the same procedural integrity checklist as was used in tutor training and was aided by a second instructor during a number of sessions. The second instructor also provided reinforcement checks and monitored procedural integrity, as well as conducting interobserver agreement. Procedural integrity was defined as the number of steps of instruction completed correctly by the tutor. Procedural integrity checklists were completed by the experimenter and second instructor during each session. The checklists provided a task analysis of the tutoring procedure and were used to demonstrate that the students were correctly completing each step of the tutoring procedures. Each step that was observed as being completed correctly by the students was marked with a checkmark. The number of steps completed correctly during each timing was then calculated.

Interobserver agreement on timings was included as part of the procedural integrity procedure. Procedural integrity was recorded for a total of 87 timings during the intervention. The average procedural integrity across tutors was 91.19%, with a range of 81.25-100%. Interobserver agreement was also carried out on 5.7% of the procedural integrity checklists carried out by the experimenter. The second instructor observed and recorded the number of steps completed correctly by the tutors and percentage of agreement on the procedural integrity on each timing was calculated. Interobserver agreement for procedural integrity was 92.73%.

Social Validity

A brief questionnaire was developed by the experimenter to assess social validity. This was administered to all the tutors and tutees who had participated following the completion of the intervention. The questionnaire combined questions based on Likert scales

of agreement and open-ended questions to gain an insight into participants' experience of the intervention. The questionnaire consisted of eight questions, the first three of which were answered using 5-point Likert scale from "strongly disagree" to "strongly agree". These questions addressed whether participants enjoyed the intervention, whether they would participate again given the opportunity, and if they thought the intervention had been helpful for them. The last 5 questions were open-ended to allow for elaboration on participants' experience. These questions addressed what participants liked most and least about the project, if they thought switching partners was beneficial, suggestions for how they might change the intervention, and any other comments they would like to make.

Data Analysis

A multivariate analysis of covariance (MANCOVA) was conducted to investigate the differences between the three groups on mathematics dependent measures at post-testing. Participants' pre-test scores were used as the covariates in the analyses. A separate MANCOVA was conducted to investigate differences in post-test scores between groups on the SSIS-RS (Gresham & Elliott, 2008). Preliminary checks were conducted for each MANCOVA to ensure there was no violation of assumptions. The assumptions tested for were univariate and multivariate outliers, normality, multicollinearity, equality of covariance matrices, homogeneity of variance, and homogeneity of regression slopes.

Results

The Experimental Group completed an average of 15.3 intervention sessions, achieving an average of 6.6 fluency aims. The number of fluency aims reached ranged from two (achieved by four students) to 18 (achieved by one student). Table 1 outlines the number of sessions completed and number of fluency aims achieved by the Experimental Group.

Insert Table 1 here

Mathematics Dependent Measures

Preliminary checks on the data indicated that the pre-test scores for fluency with target math facts violated the assumption of normality, $p = .03$. The data was transformed using a log transformation. Outliers were identified in the post-test of fluency with target math facts and the WJIII (Woodcock, McGrew, & Mather, 2001b) Mathematics Fluency subtest; therefore, transformations were also carried out on these variables. The transformed variables were subsequently included in the MANCOVA. The preliminary checks found that there were no other violations of assumptions.

The results of the MANCOVA indicated a statistically significant difference between groups on the combined dependent variable, $\lambda = .6$, $F(6, 66) = 3.198$, $p = .008$, $\eta_p^2 = .225$. Multivariate tests were followed up with univariate analyses of covariances (ANCOVA) to investigate the effect of intervention, controlling for pre-tests, on each of the dependent variables separately. ANCOVA results demonstrated that there were no significant differences between groups for scores on the Calculation subtest of the WJIII (Woodcock, McGrew, & Mather, 2001b), $F(2, 35) = 1.721$, $p = .194$, $\eta_p^2 = .090$. Significant differences between groups were found for the Mathematics Fluency subtest of the WJIII (Woodcock, McGrew, & Mather, 2001b), $F(2, 35) = 5.398$, $p = .009$, $\eta_p^2 = .236$, and for fluency with target math facts, $F(2, 35) = 8.504$, $p = .001$, $\eta_p^2 = .327$.

Bonferroni post hoc pairwise comparisons were conducted to assess pairwise differences in adjusted group means (controlling for pre-tests as covariates). The Experimental Group's adjusted means were significantly higher than the Control Group for both the WJIII (Woodcock, McGrew, & Mather, 2001b) Mathematics Fluency subtest ($p = .014$) and fluency with target math facts ($p = .001$). Results demonstrated no significant differences for means on any of the mathematics dependent measures between the Tutoring

Group and either the Control or Experimental Group. Table 2 shows the original mean scores for each group on pre- and post-test mathematics dependent measures.

Insert Table 2 here

SSIS-RS

MANCOVA was conducted to investigate the differences between the three groups on social skills and problem behaviors (SSIS-RS; Gresham & Elliott, 2008) at post-testing. Four participants were excluded from the MANCOVA due to incomplete responses. The remaining 37 participants were included in the analysis. Preliminary checks on the data indicated that the post-test scores for the Problem Behavior subscale violated the assumption of normality, $p = .005$. The data were transformed using an inverse transformation and the transformed variables were included in the MANCOVA analysis. The preliminary checks found that there were no other violations of assumptions. The results of the MANCOVA found that there was no statistically significant difference between groups on the combined dependent variables, measuring social skills and problem behaviors, $\lambda = .95$, $F(4, 62) = .45$, $p = .776$, $\eta_p^2 = .03$. Table 2 shows the original mean scores for each group on the SSIS-RS (Gresham & Elliott, 2008).

Social Validity

The social validity measure provided an account of participants' attitudes towards their participation in the intervention. The vast majority of participants enjoyed being part of the intervention, with 96% of the participants indicating that they liked taking part. Ninety-one percent of tutors and 80% of tutees said they would participate in the intervention again if they could. Ninety-one percent of tutors and 100% of tutees felt that the intervention had helped them with their mathematics. Sixty-seven percent of tutors and 80% of tutees preferred rotating the dyad pairs to remaining in the same pairing for the full intervention. Tutors indicated that their favourite aspects of intervention included; 'helping others get their

goals’ and ‘teaching the tutees’, while most of the tutees stated that their favourite aspect was either that it helped them with their mathematics or the reinforcement aspect of the intervention. Responses from the tutees mainly indicated that there was ‘nothing’ they liked least about the intervention, while some of those participating as tutors highlighted the behaviors of their tutoring partners as the thing they liked least about the intervention ‘doing maths with people that don’t listen’ and ‘that some people cheated to go onto the next set’.

Discussion

The current study evaluated the outcomes of a school-based fluency intervention incorporating cross-age peer tutoring with frequency-building using SAFMEDS and targets from the Morningside Math Facts: Addition and Subtraction curriculum (Johnson, 2008). The intervention targeted fluency with basic mathematics component skills and employed an RCT design to evaluate its effectiveness. The Experimental Group demonstrated the ability to progress through the intervention with many achieving a considerable number of fluency aims during the 8-week intervention. Significant differences on mean post-test scores between the Experimental and Control Groups were found for the Mathematics Fluency subtest of the WJIII (Woodcock, McGrew, & Mather, 2001b) and on fluency with target math facts. There were no significant differences between groups on the Calculation subtest of the WJIII (Woodcock, McGrew, & Mather, 2001b), or on measures of social skills and competing problem behaviors (SSIS-RS; Gresham & Elliott, 2008).

The current findings are in agreement with previous research showing the effectiveness of cross-age peer tutoring (Hawkins et al., 2009; Robinson et al., 2005; Fueyo & Bushell, 1998) and fluency-based interventions (Gallagher, 2006; Hartnedy et al., 2005; Mc Tiernan et al., 2016; Poncy et al., 2010) to increase fluency with mathematics skills. Cross-age peer tutoring has not been as widely researched in the literature as homogeneous peer-tutoring and has often included participants with learning disabilities or emotional and

behavioral difficulties (Heron et al., 2003; Rohrbeck et al., 2003). The current study supports the use of cross-age peer tutoring with typically developing students from a disadvantaged population. It provides an important addition to the literature on cross-age peer tutoring for mathematics, demonstrating that it provides an effective intervention for younger tutees. This has further important implications for practice, indicating that cross-age peer tutoring is an instructional approach that would afford beneficial outcomes if incorporated into educational settings. The relatively low demands on teachers further supports cross-age peer tutoring as a favourable strategy (Dufrene et al., 2010).

The current study provides evidence for an instructional approach, utilising the combination of peer tutoring and frequency-building using SAFMEDS, to increase rates of correct responding with math fact families. Further, most participants reported that they enjoyed taking part in this instructional approach. The implications for educational settings are significant, given the importance of math fact mastery in the progression to more complex skills and overall mathematical ability (Carr et al., 2008; Jordan et al., 2003b; Nelson et al., 2013) and the significant correlation which exists between SES and mathematics performance (OECD, 2012). The current study provides evidence for the efficacy of this approach as well as demonstrating the ease with which such an approach can be utilised for an elementary-school population.

The findings also contribute to the literature base supporting fluency-based interventions to improve performance with component mathematics skills. The successful use of fluency-based interventions has been demonstrated within the literature for a number of years, yet there is still a lack of empirically validated approaches (Coddington et al., 2009). Few studies to date have used RCTs to evaluate such approaches meaning that the current study contributes significantly in establishing evidence-based instructional practices. Employing an RCT within an educational setting promotes confidence that the treatment

outcomes reflect an empirically robust investigation of the fluency intervention. Though often considered impractical to establish, RCTs within fluency research will further support educational applications for children from disadvantaged backgrounds.

There were no significant differences between the Tutor Group and Control Group across mathematics dependent measures. The use of SAFMEDS as a method of instruction for math facts could potentially have impacted on the lack of academic gains demonstrated by the Tutor Group. The benefits observed for tutors in previous research is often attributed to the process of constructing explanations for tutees on the material being taught, requiring the tutor to elaborate and generate connections between the current material and previously learned information (Fuchs et al., 2002). For example, in an intervention involving teaching of algebra, tutors combine their knowledge of basic math facts with the algebraic equations they are helping their tutees to learn. As the answers to the math facts were provided on the reverse of the cards, there was little opportunity for learning for the tutors, who could simply read the answers to math facts from the cards. Future research should evaluate possible benefits for peer tutors using different methods of instruction and differing target skills.

The tutoring intervention did not significantly impact measures of social skills or problem behaviors of either tutors or tutees. This is inconsistent with previous research demonstrating that academic tutoring has beneficial collateral effects on social outcomes (Robinson et al., 2005). The variability in scores on the SSIS-RS at pre- and post- testing indicates that this measure may not have been representative of true levels of social skills. Individual scores varied widely from pre- to post-tests, with both increases and decreases in each scale across the groups. There was no discernible trend for either the Social Skills or Problem Behaviors in any of the three groups. An alternative measure, that was not reliant on subjective reports, may have been more suitable to determine a more accurate picture of participants' social skills and problem behaviors.

Limitations and Future Directions

As mentioned above, the variability in scores on the SSIS-RS (Gresham & Elliott, 2008) indicates that it may not be a sensitive enough measure for assessing social skills and problem behaviors. Subjective measures have been found in previous research to show no correlation with independent, objective measures related to the variable of interest, and can be difficult to interpret because they are often expressed in ordinal scales (Jahedi & Mendez, 2013). Future research should also include objective measures, such as direct observation of behavior, or include additional subjective measures, such as teacher ratings in addition to the self-ratings used in the current study. These methods would provide a more sensitive measure of participant behaviors and social skills, rather than relying on participant's own reports.

The level of interobserver agreement collected for pre- and post-test fluency probes is a limitation to the current study, as collection for both was less than that which is traditionally recommended. In addition, interobserver agreement was not collected for the WJIII subtests (Woodcock, McGrew, & Mather, 2001b) and the SSIS-RS (Gresham & Elliott, 2008) administered pre- and post-test. Future studies should ensure that sufficient levels of IOA are collected for all dependent measures to improve reliability of measurement.

Participants achieved a wide range of fluency aims across the 8 weeks of intervention, ranging from 2 to 18, despite similar number of sessions. Intervention data was charted on a Standard Celeration Chart to monitor each participant and to ensure that they were making progress. However, some participants achieved only two fluency aims. Some participants most likely required more time to achieve each fluency aims and because the intervention was of a relatively low intensity and of short duration, did not achieve more during the study. However, it may have been beneficial to add criteria for modifications to instruction (e.g.,

reversing to a previous fluency aim) if participants did not reach a particular aim within a set number of sessions as demonstrated on the Standard Celeration Charts.

Follow-up assessments of retention should also be included in future studies to determine if the short-term gains observed during intervention have been maintained. Due to time constraints imposed, it was not possible to conduct such a follow-up assessment.

Retention is one of the critical outcomes associated with the fluent performance of skills (Weiss, 2001). It would be beneficial to determine if the short term academic gains translated to long term gains over time and if continued instruction with component skills would lead to improvements in other Mathematics applications over time.

Conclusion

Proficiency in basic computational skills is essential for independent adaptive functioning in numerous areas of daily life. Despite this, there is a widely acknowledged deficit in students' performance of such skills and a scarcity of research on effective, empirically-based interventions to increase fluency in these skills (Coddington, Archer, & Connell, 2010; Coddington et al., 2009; Poncy, Fontenelle, & Skinner, 2013). The current study provides support for the incorporation of cross-age peer tutoring and frequency-building using SAFMEDS to increase fluency with math fact families. Relatively little of this research has focused on cross-age peer tutoring for a mainstream disadvantaged population while a lack of sufficient research on empirically-validated instructional approaches to increase fluency also exists in the literature. This study aimed to address such gaps in the literature and to assess the benefits of such an intervention for both tutees and tutors. The findings suggest that, while there may be limited benefits for tutors, the use of cross-age peer tutoring and fluency-based instruction has significantly benefits for younger tutees.

Compliance with Ethical Standards

Ethical approval. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent. Informed consent was obtained from all individual participants included in the study.

References

- Axtell, P. K., McCallum, R. S., Bell, S. M., & Poncy, B. (2009). Developing math automaticity using a classwide fluency building procedure for middle school students: A preliminary study. *Psychology in the Schools, 46*, 526-538. doi: 10.1002/pits.20395
- Beck, R., & Clement, R. (1991). The Great Falls Precision Teaching project: A historical examination. *Journal of Precision Teaching, 8*, 8-12.
- Beverly, M., Hughes, J. C., Hastings, R. P. (2009). What's the probability of that? Using SAFMEDS to increase undergraduate success with statistical concepts. *European Journal of Behavior Analysis, 10*, 183-195. doi:
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163-197. doi: 10.4135/9781412950534.n3022
- Bliss, S. L., Skinner, C. H., McCallum, E., Saecker, L. B., Rowland-Bryant, E., & Brown, K. S. (2010). A comparison of taped problems with and without a brief post-treatment assessment on multiplication fluency. *Journal of Behavioral Education, 19*, 156-168. doi: 10.1007/s10864-010-9106-5
- Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review, 43*, 260-285.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R.

(2013). Academic Benefits of Peer Tutoring: A Meta-Analytic Review of Single-Case Research. *School Psychology Review* (1), 39-55.

Brady, N. C. (1997). The teaching game: A reciprocal peer tutoring program for preschool children. *Education and Treatment of Children*, 20, 123-149.

Brady, K. K., & Kubina, R. M. (2010). Endurance of multiplication fact fluency for students with attention deficit hyperactivity disorder. *Behavior Modification*, 34, 79-93. doi: 10.1177/0145445510361331

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly*, 13, 140-163. doi: 10.1111/j.1937-8327.2000.tb00180.x

Burns, M. K. (2005). Using incremental rehearsal to practice multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children*, 28, 237-249.

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology*, 103(3), 617. doi: 10.1037/a0023864

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences*, 18, 61-75. doi: 10.1016/j.lindif.2007.04.005

Casey, J., McLaughlin, T. F., & Everson, M. (2003). The effects of five minute practice, unlimited practice, with SAFMED cards on correct and error rate in math facts for two elementary school children with learning disabilities. *International Journal of Special Education*, 18, 66-72.

- Cates, G. L., & Rhymer, K. N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education, 12*, 23-34. doi: 10.1023/A:1022318321416
- Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension of using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioral Education, 19*, 93-105. doi: 10.1007/s10864-010-9102-9
- Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*(1), 36-47. doi: 10.1111/j.1540-5826.2010.00323.x
- Codding, R. S., Hilt-Panahon, A., Panahon, C. J., & Benson, J. L. (2009). Addressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education and Treatment of Children, 32*, 279-312. doi: 10.1353/etc.0.0053
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248. doi: 10.3102/00028312019002237
- Daly, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36*(4), 562-581.
- Department of Education and Science. (2005). *DEIS (Delivering Equality of Opportunities in Schools) An Action Plan for Educational Inclusion*. Retrieved from: https://www.education.ie/en/Publications/PolicyReports/deis_action_plan_on_educati educ_inclusion.pdf

- Dufrene, B. A., Noell, G. H., Gilbertson, D. N., & Duhon, G. J. (2005). Monitoring implementation of reciprocal peer tutoring: Identifying and intervening with students who do not maintain accurate implementation. *School Psychology Review, 34*, 74-86.
- Dufrene, B. A., Reisener, C. D., Olmi, D. J., Zoder-Martell, K., McNutt, M. R., & Horn, D. R. (2010). Peer tutoring for reading fluency as a feasible and effective alternative in response to intervention systems. *Journal of Behavioral Education, 19*, 239-256. doi: 10.1007/a10864-101-9111-8.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behavior Research Methods, 39*, 175-191 doi: 10.3758/BF03193146
- Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal, 101*, 495-510. doi: 10.1086/499684
- Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31*, 79-92. Doi: 10.2307/20528819
- Fuchs, L. S., Fuchs, D., Yazdian, L., & Powell, S. R. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review, 31*, 569-583.
- Fueyo, V., & Bushell, D. (1998). Using number line procedures and peer tutoring to improve the mathematics computation of low-performing first graders. *Journal of Applied Behavior Analysis, 31*, 417-430. doi: 10.1901/jaba.1998.31-417
- Gallagher, E. (2006). Improving a mathematical key skill using precision teaching. *Irish Educational Studies, 25*, 303-319. doi: 10.1080/03323310600913757

- Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science*, 22, 23-27. doi: 10.1177/0963721412469398
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, 104, 206-223. doi: 10.1037/a0025398
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, 33, 18-28. doi: 10.1177/002246699903300102
- Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, 98, 732-749. doi: 10.1037/0022-0663.98.4.732
- Graf, S. A., & Lindsley, O. R. (2002). *Standard Celeration Charting 2002*. Poland: OH: Graf Implements.
- Gresham, F.M., & Elliott, S.N. (1990). *Social skills rating system*. Minneapolis, MN:NCS Pearson.
- Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales manual*. Minneapolis, MN: NCS Pearson.
- Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24–49). Cambridge, MA: MIT Press.
- Gross, T. J., Duhon, G J., Hansen, B., Rowland, J. E., Schutte, G., & Williams, J. (2014). The effect of goal-line presentation and goal selection on first-grader subtraction fluency.

The Journal of Experimental Education, 82, 555-571. doi:

10.1080/00220973.2013.813369

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23-40). Columbus: Charles E. Merrill Publishing Co.

Hartnedy, S. L., Mozzoni, M. P., & Fahoum, Y. (2005). The effect of fluency training on math and reading skills in neuropsychiatric diagnosis children: A multiple baseline design. *Behavioral Interventions*, 20(1), 27-36. doi: 10.1002/bin.167

Hawkins, R. O., Musti-Rao, S., Hughes, C., Berry, L., & McGuire, S. (2009). Applying a randomized interdependent group contingency component to classwide peer tutoring for multiplication fact fluency. *Journal of Behavioral Education*, 18, 300-318. doi: 10.1007/s10864-009-9093-6

Heron, T. E., Welsch, R. G., & Goddard, Y. L. (2003). Applications of tutoring systems in specialized subject areas: An analysis of skills, methodologies, and results. *Remedial and Special Education*, 24, 288-300. doi: 10.1177/07419325030240050401

Hughes, J. C., Beverly, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis*, 8, 221-238.

Jacobson, J., Thrope, L., Fisher, D., Lapp, D., Frey, N., & Flood, J. (2001). Cross-age tutoring: A literacy improvement approach for struggling adolescent readers. *Journal of Adolescent & Adult Literacy*, 44, 528-536.

Jahedi, S., & Mendez, F. (2013). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior & Organisation*, 98, 97-114. doi: 10.1016/j.jebo.2013.12.016

- Johnson, K. (2008). *Morningside Mathematics Fluency: Math Facts* (Vols. 1-6; curriculum program). Seattle, WA: Morningside Press.
- Johnson, K., & Street, E. M. (2013). *Response to intervention and precision teaching*. New York: The Guilford Press.
- Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834–850. doi: 10.1111/1467-8624.00571
- Kena, G, Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., ... Dunlop Velez, E. (2016). *The Condition of Education 2016* (NCES, 2016-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from: <http://nces.ed.gov/pubs2016/2016144.pdf>
- Kubina Jr, R. M., Yurich, K. K., Durica, K. C., & Healy, N. M. (2016). Developing Behavioral Fluency with Movement Cycles Using SAFMEDS. *Journal of Behavioral Education, 25*(1), 120-141.
- Kubina, R. M., & Wolfe, P. (2005). Potential applications of behavioral fluency for students with autism. *Exceptionality, 13*, 35-44. doi: 10.1207/s15327035ex1301_5
- Mc Tiernan, A., Holloway, J., Healy, O., & Hogan, M. (2016). A randomized controlled trial of the Morningside math facts curriculum on fluency, stability, endurance and application outcomes, *Journal of Behavioral Education, 24*, 1-20. doi: 10.1007/s10864-015-9227-y
- Meindl, J. N., Ivy, J. W., Miller, N., Neef, N. A., & Williamson, R. L. (2013). An examination of stimulus control in fluency-based strategies: SAFMEDS and generalization. *Journal of Behavioral Education, 22*, 229-252. doi: 10.1007/s10864-013-9172-6

- Miller, A. D., Hall, S. W., Heward, W. L. (1995). Effects of sequential 1-minute time trials with and without inter-trial feedback and self-correction on general and special education students' fluency with math facts. *Journal of Behavioral Education*, 5, 319-345. doi: 10.1007/BF02110318
- Mitchell, R. J., Morrison, T. G., Feinauer, E., Wilcox, B., & Black, S. (2015). Effects of fourth and second graders' cross-age tutoring on students' spelling. *Reading Psychology*, 0, 1-20. doi: 10.1080/02702711.2015.1025164
- Mong, M. D., & Mong, K. W. (2010). Efficacy of two mathematics interventions for enhancing fluency with elementary students. *Journal of Behavioral Education*, 19, 273-288. doi: 10.1007/s10864-010-9114-5
- Nelson, P. M., Burns, M. K., Kanive, R., & Ysseldyke, J. E. (2013). Comparison of a math fact rehearsal and a mnemonic strategy approach for improving math fact fluency. *Journal of School Psychology*, 51, 659-667. doi:10.1016/j.jsp.2013.08.003
- Nobel, M. N. (2005). *Effects of classwide peer tutoring on the acquisition, maintenance, and generalization of science vocabulary words for seventh grade students with learning disabilities and/or low achievement* (Doctoral dissertation, The Ohio State University). Retrieved from <https://etd.ohiolink.edu/>
- OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD Publishing, Paris. doi:10.1787/9789264266490-en
- Patton, J. R., Cronin, M. E., Bassett, D. S., & Koppel, A. E. (1997). A life skills approach to mathematics instruction preparing students with learning disabilities for the real-life math demands of adulthood. *Journal of Learning Disabilities*, 30(2), 178-187. doi: 10.1177/002221949703000205
- Petrill, S., Logan, J., Hart, S., Vincent, P., Thompson, L., Kovas, Y., & Plomin, R. (2012). Math fluency is etiologically distinct from untimed math performance, decoding

- fluency, and untimed reading performance: Evidence from a twin study. *Journal of Learning Disabilities*, 45, 371-381. doi: 10.1177/0022219411407926
- Poncy, B. C., Duhon, G. J., Lee, S. B., & Key, A. (2010). Evaluation of techniques to promote generalization with basic math fact skills. *Journal of Behavioral Education*, 19(1), 76-92. doi: 10.1007/s10864-010-9101-x
- Poncy, B. C., Fontenelle IV, S. F., & Skinner, C. H. (2013). Using Detect, Practice, and Repair (DPR) to Differentiate and Individualize Math Fact Instruction in a Class Wide Setting. *Journal of Behavioral Education*, 22(3), 211-228. doi: 10.1007/s10864-013-9171-7
- Rivera-Batiz, F. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, 27, 313-328. doi: 10.2307/145737
- Robinson, D. R., Schofield, J. W., & Steers-Wentzell, K. L. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review*, 17, 327-362. doi: 10.1007/s10648-005-8137-2
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95, 240-257. doi: 10.1037/0022-0663.95.2.240
- Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical Abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing.
- Schutte, G. M., Duhon, G. J., Solomon, B. G., Poncy, B. C., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice basic math fact fluency growth rates. *Journal of School Psychology*, 53, 149-159. doi: 10.1016/j.jsp.2014.12.003

- Shiel, G., Kavanagh, L., & Millar, D. (2014). *The national assessments of English reading and mathematics, volume 1: Performance report*. Dublin: Educational Research Centre.
- Skinner, C. H. (2002). An empirical analysis of interspersal research: Evidence, implications, and applications of the discrete task completion hypothesis. *Journal of School Psychology, 40*(4), 347-368. doi: 10.1016/s0022-4405(02)00101-2
- Smyth, E., & McCoy, S. (2009). *Investing in education: Combating educational disadvantage*. Dublin: ESRI Research Series No. 6
- Sotelo-Dynega, M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. F. (2013). English language proficiency and test performance: An evaluation of bilingual students with the Woodcock-Johnson III tests of cognitive abilities. *Psychology in the Schools, 50*(8), 781-797. doi: 10.1002/pits.21706
- Sprinthall, N. A., & Scott, J. R. (1989). Promoting psychological development, math achievement, and success attribution of female students through deliberate psychological education. *Journal of Counselling Psychology, 36*, 440-446. doi: 10.1037//0022-0167.36.4.440
- Stenhoff, D. M., & Lignusaris-Kraft, B. (2007). A review of the effects of peer tutoring on students with mild disabilities in secondary settings. *Exceptional Children, 74*(1), 8-30. doi: 10.1177/001440290707400101
- Surgenor, P., Shiel, G., Close, S., & Millar, D. (2006). *Counting on success: Mathematics achievement in Irish primary schools*. Dublin: Department of Education and Science.
- Thompson, L. A., & Kelly-Vance, L. (2001). The impact of mentoring on academic achievement of at-risk youth. *Children and Youth Services Review, 23*, 227-242. doi: 10.1016/s0190-7409(01)00134-7
- Topping, K.J. (1988). *The peer tutoring handbook: Promoting cooperative learning*. London,

UK: Croom Helm.

Topping, K. J., Miller, D., Murray, P., Henderson, S., Fortuna, C., & Conlin, N. (2011).

Outcomes in a randomised controlled trial of mathematics tutoring. *Educational Research*, 53, 51-63. doi: 10.1080/00131881.2011.552239

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., & Miller, D. (2011). Improving

attainment across a whole district: School reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 22, 265-289, doi:10.1080/09243453.2011.589859

Houten, R. V., & Thompson, C. (1976). The Effects of Explicit Timing on Math

Performance. *Journal of Applied Behavior Analysis*, 9(2), 227-230.

Weiss, M. J. (2001). Expanding aba intervention in intensive programs for children with

autism: The inclusion of natural environment training and fluency based instruction. *The Behavior Analyst Today*, 2, 182-186. doi: 10.1037/h0099946

Weiss, M. J., Pearson, N., Foley, K., & Pahl, S. (2010). The importance of fluency outcomes

in learners with autism. *The Behavior Analyst Today*, 11, 245-252. doi:10.1037/h0100704

Woodcock, R., Mc Grew, K. S., & Mather, N. (2001b). *Woodcock-Johnson Tests of*

Achievement, Third Edition. Illinois: The Riverside Publishing Company.